Spatial Statistical Software

by

Stephen L. Rathbun
Department of Health Administration, Biostatistics, and Epidemiology
College of Public Health
University of Georgia
Athens, GA 30605
rathbun@uga.edu

1

General Observations:

- Perhaps the single most limiting factor for dissemination a modern spatial statistical procedures is the limited availability of statistical software.
- Writing of statistical software involves the following trade-off:
  - Ease of use
  - Flexibility

2

Outline: Review software for three areas of spatial statistics.

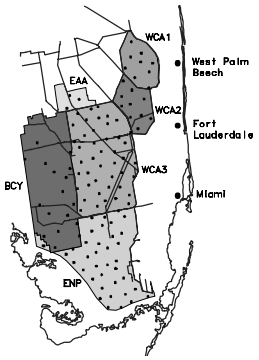
1. Geostatistics.
2. Spatial Point Patterns.
3. Lattice Data.

3

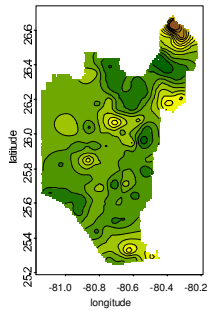
Geostatistics

South Florida Ecosystem Assessment.

Sample Sites



Predicted Total Mercury



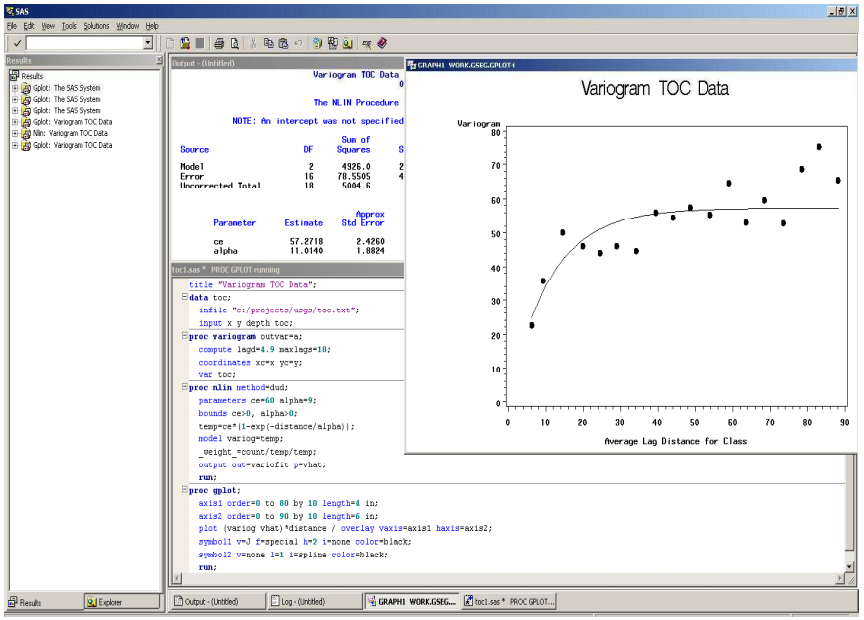
4

Geostatistical Software:

- SAS
- Surfer
- ArcGIS Geostatistical Analyst
- S+SpatialStats
- R

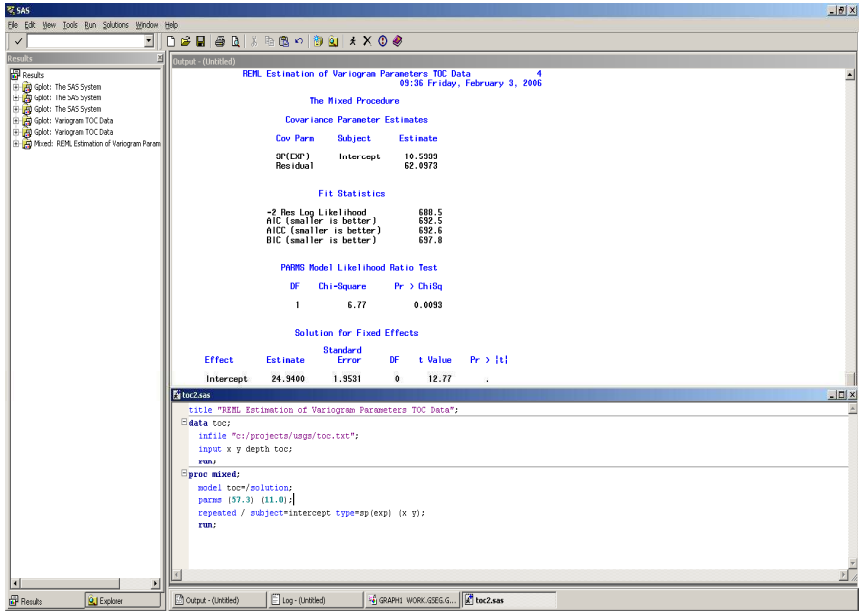
5

SAS: Variogram Model Fitting



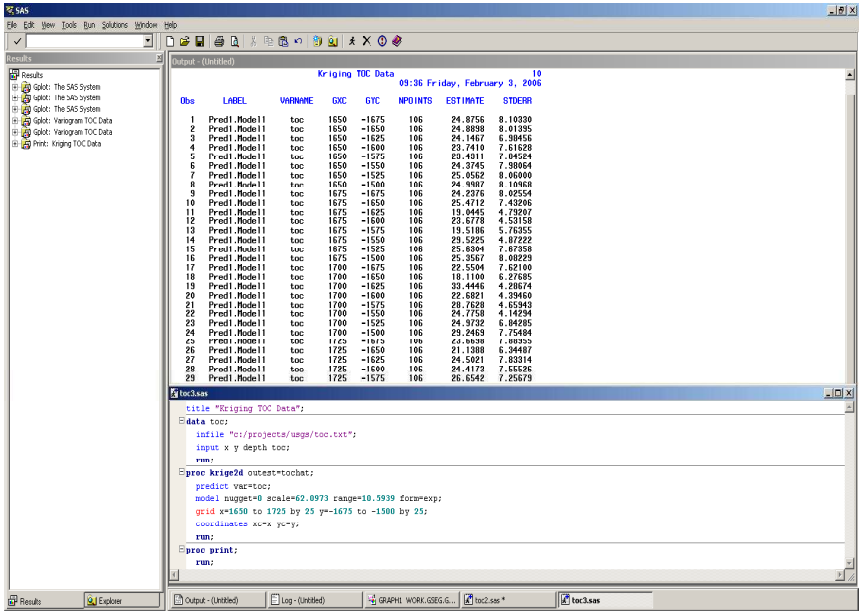
6

SAS: REML Estimation of Variogram Model Parameters



7

SAS: Kriging



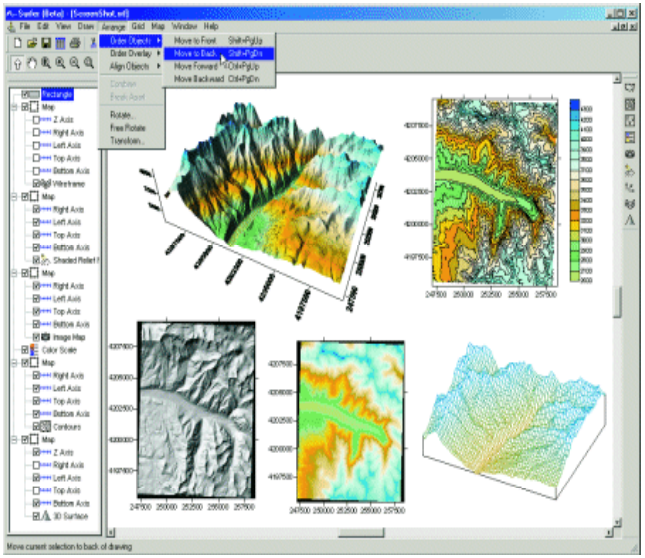
8

Comments: SAS Geostatistics

- SAS is not menu driven. Analysis is carried out by writing SAS programs in the SAS editor.
  - For those who have experience with SAS, the geostatistical procedures are easy to apply.
  - Harder to use than menu-driven software.
- SAS has procedures for:
  - Isotropic and anisotropic variogram estimation (proc variog);
  - Variogram model fitting:
    - Weighted Least Squares (proc nlin);
    - Maximum Likelihood and REML (proc mixed).
  - Ordinary Kriging (proc krige2d).
  - Universal Kriging (proc mixed).
  - Generalized Mixed Models (proc glimmix).
- Limitations:
  - Limited choice of variogram models.
  - Cannot draw good contour maps.

9

Surfer
http://www.goldensoftware.com/



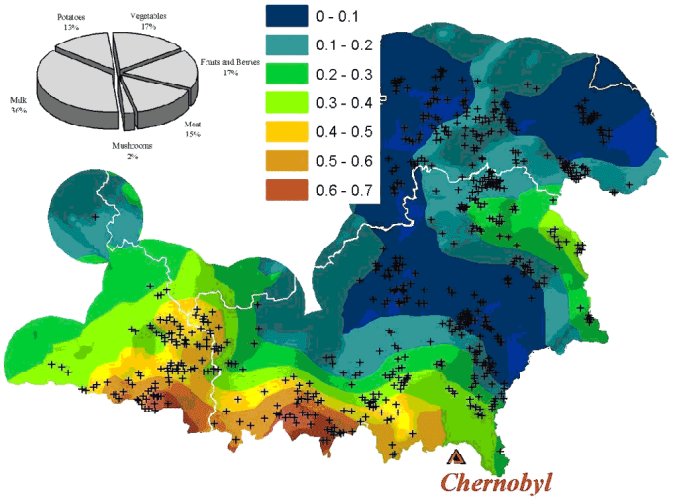
10

Comments: Surfer

- Menu Driven
- Surfer has procedures for:
  - Variogram Estimation;
  - Least Squares Estimation of Variogram Model Parameters
  - Wide Variety of Variogram Models: exponential, Gaussian, linear, log, power, quadratic, rational quadratic, spherical, wave, pentaspherical, cubic.
  - Ordinary Kriging
  - Excellent mapping capabilities: contour maps; 3D surface maps; wireframe maps; vector maps; shaded relief maps.
- Limitations:
  - Cannot fit Matern variogram;
  - Universal kriging not available.

11

ArcGIS Geostatistical Analyst
http://www.esri.com/software/arcgis/extensions/geostatistical/index



12

Comments: ArcGIS Geostatistical Analyst

- Menu Driven
- Geostatistical Analyst has procedures for:
  - Isotropic and anisotropic variogram estimation;
  - Least squares estimation of variogram parameters;
  - Wide variety of variogram models: circular, spherical, tetraspherical, pentaspherical, exponential, Gaussian, rational quadratic, hole effect, k-bessel, stable.
  - Variety of kriging methods:
    - Ordinary kriging
    - Universal kriging
    - Indicator kriging (Binary Variables)
    - Disjunctive kriging (Nonliner Geostatistics)
    - Cokriging (Multivariate Geostatistics)
  - Crossvalidation for model diagnostics.
- Limitation: Expensive ($2,500 for Geostatistical Analyst, $1,500 for ArcView 9.1)

Definition: Crossvalidation.

- Remove the data at site $\mathbf{s}_i$ from the data set;
  - Use the remaining data to obtain the kriging predictor $\hat{Z}_{-i}(\mathbf{s}_i)$ of the data at site $\mathbf{s}_i$
  - Compute the corresponding kriging variance $\sigma^2_{-i}(\mathbf{s}_i)$
- Repeat the above procedure for all sites.
- Compare observed values $Z(\mathbf{s}_i)$ with predicted values $\hat{Z}_{-i}(\mathbf{s}_i)$
  - Bias Measure

$$\mathrm{CV}_1 = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{Z(\mathbf{s}_i) - \hat{Z}_{-i}(\mathbf{s}_i)}{\sigma_{-i}(\mathbf{s}_i)}\right\}$$

  - Uncertainty Assessment

$$\mathrm{CV}_2 = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{Z(\mathbf{s}_i) - \hat{Z}_{-i}(\mathbf{s}_i)}{\sigma_{-i}(\mathbf{s}_i)}\right\}^2$$

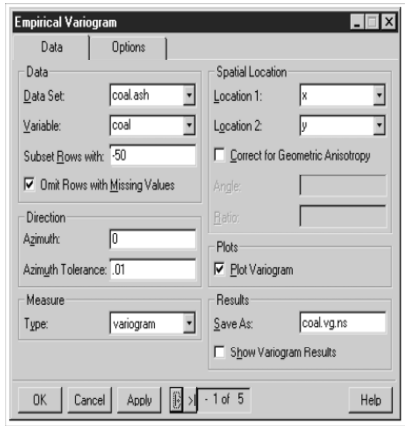
For a valid model, we should have

$$\mathrm{CV}_1 \cong 0 \text{ and } \mathrm{CV}_2 \cong 1$$

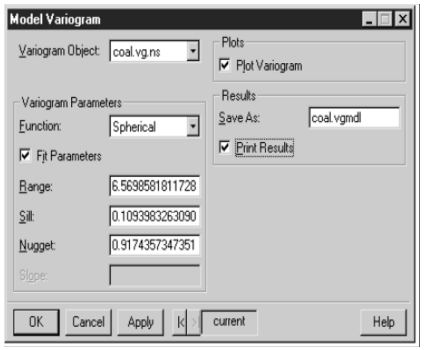
S+SPATIALSTATS

http://www.insightful.com/products/spatial/default.asp

Variogram Estimation

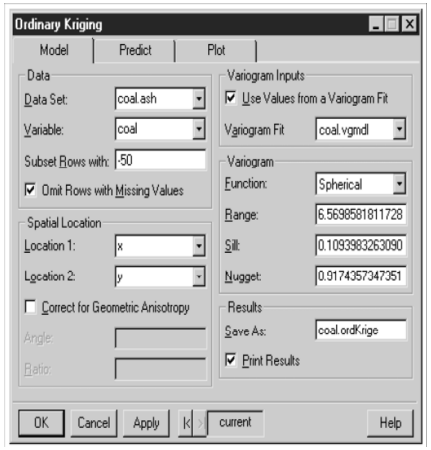


Least Squares Estimation

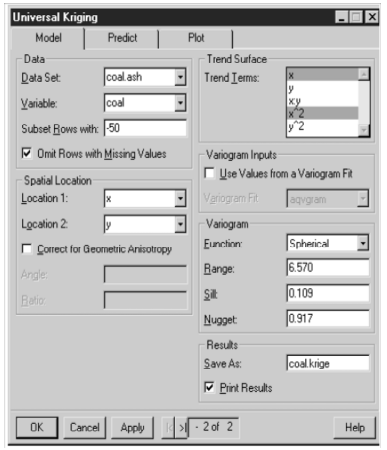


S+SPATIALSTATS

Ordinary Kriging



Universal Kriging



Comments: S+SPATIALSTATS

- Menu Driven
- S+SPATIALSTATS has procedures for:
  - Isotropic and Anisotropic Variogram Estimation
  - Least Squares Estimation of Variogram Parameters (Weighted least squares with some work)
  - Limited variogram models: Spherical, exponential, Gaussian
  - Ordinary and Universal Kriging
  - Good quality contour maps
- Software has not been kept up to date.
  - Effort has been made to improve user interface.
  - No effort has been made to include modern methods.

17

R

http://www.r-project.org/

Geostatistical Packages

- geoR http://www.est.ufpr.br/geoR/

  Frequentist and Bayesian geostatistics.
- geoRglm http://www.daimi.au.dk/~olefc/geoRglm/

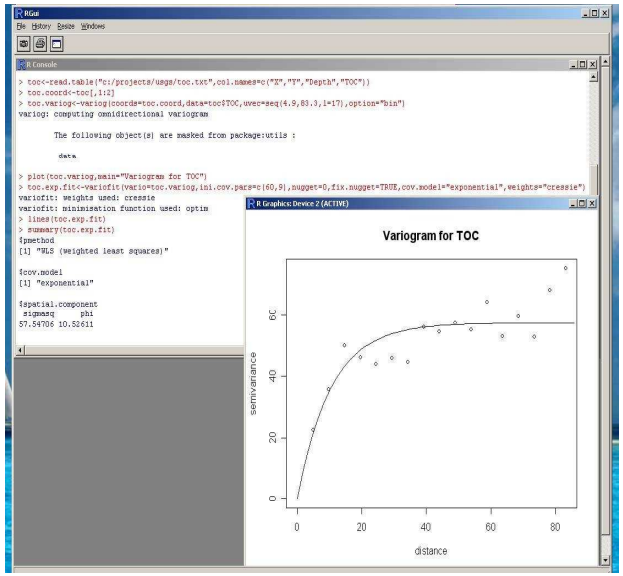
  Geostatistics for counts data. Poisson and binomial models.
- fields http://www.image.ucar.edu/GSP/Software/Fields/

  Best for global data. Includes great circle distance.
- gstat http://www.gstat.org/
- RandomFields http://www2.hsu-hh.de/schlath/R/RandomFields/RandomFields_doc.

  Spatial simulation.

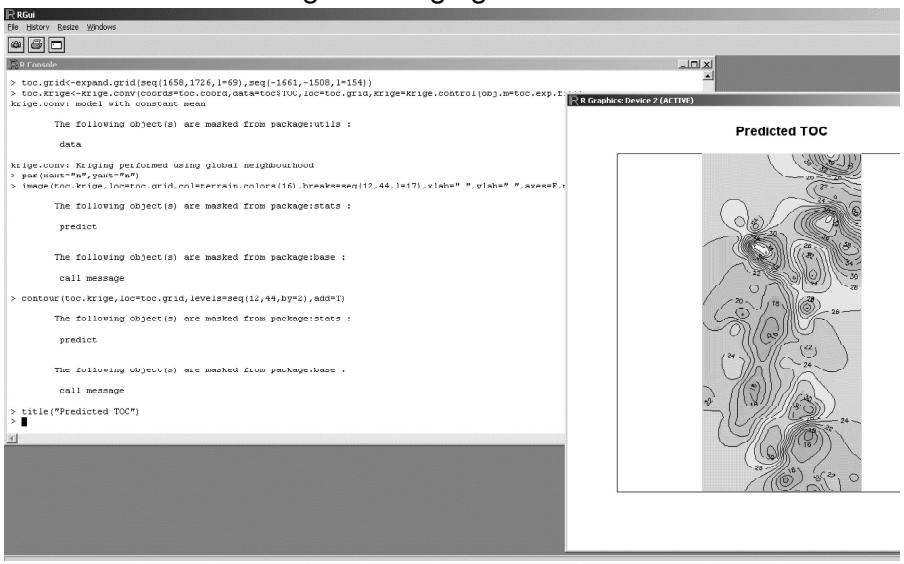
18

geoR: Variogram Modeling



19

geoR: Kriging



20

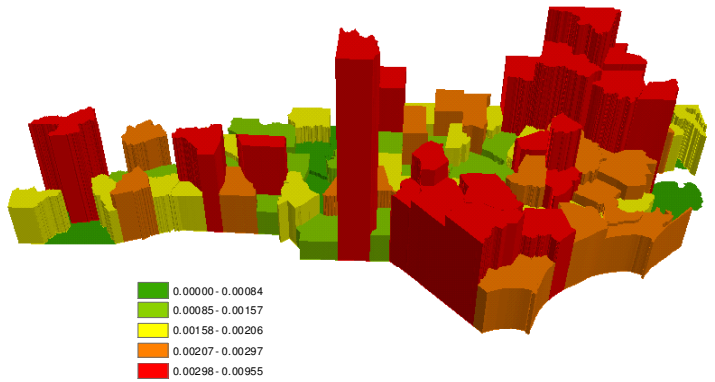
Comments: R

- Public domain software;
- Packages contributed by statistical researchers keep the software up to date;
- Command driven and interactive;
- GeoR has procedures for:
  - Variogram estimation;
  - Least squares, weighted least squares, REML estimation of variogram parameters;
  - Bayesian inference for model parameters;
  - Diverse variety of variogram models including the Matérn class;
  - Ordinary, universal and Bayesian kriging.
- GeoRglm has procedures for binomial and Poisson models for counts data;
- Fields includes great circle distance for investigating global data;
- Limitation: Not well documented.

21

Lattice Data

Sudden Infant Death Rates in North Carolina



22

Lattice Model Software

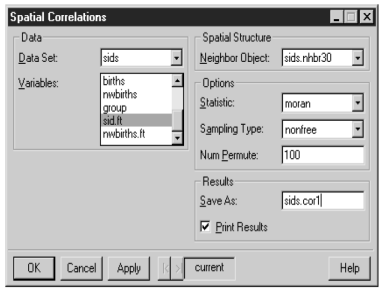
- S+SPATIALSTATS
- BUGS
- R package: spdep

23

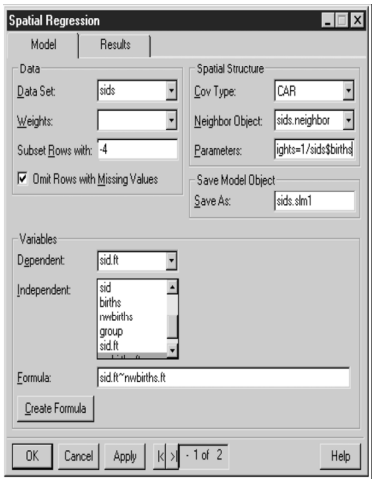
S+SPATIALSTATS

Lattice Models

Moran's Index



CAR Model



24

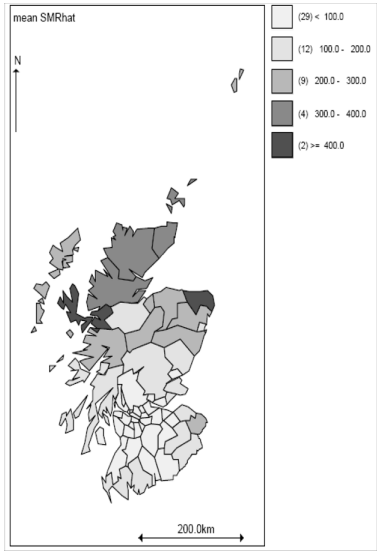
Comments: S+SPATIALSTATS

- Menu Driven
- S+SPATIALSTATS has procedures for:
  - Defining neighborhood matrices
  - Defining spatial weights matrices
  - Computing Moran's I
  - Fitting spatial regression models:
    - Conditional AutoRegressive
    - Simultaneous AutoRegressive
    - Moving Average

25

GeoBUGS

http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/geobugs.shtml



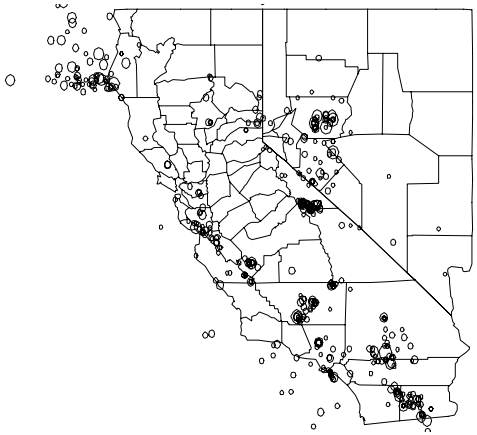
26

Comments: GeoBUGS

- Public domain software;
- Bayesian inference for lattice models:
  - CAR models
  - Poisson and binomial models with spatially dependent random effects.
- Data interface can use some work.

27

Spatial Point Pattern
California Earthquakes



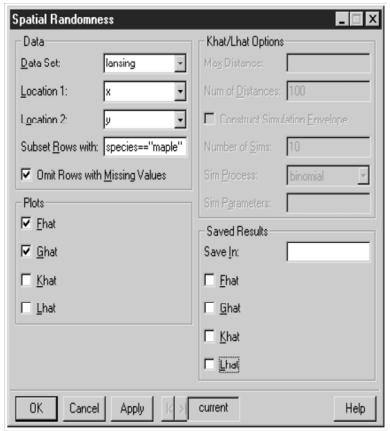
28

Point Pattern Software:

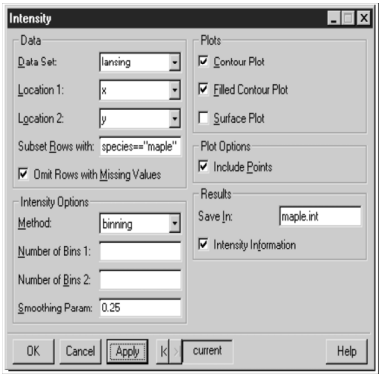
- S+SPATIALSTATS
- R

S+SPATIALSTATS

K-Function



Nonparametric Intensity



Comments: S+SPATIALSTATS

- Menu Driven;
- S+SPATIALSTATS has procedures for:
  - Computing F-, G- and K-functions;
  - Testing complete spatial randomness;
  - Nonparametric estimation of the intensity function;
  - Fitting the point cluster process model.

R: **Point Pattern Packages**:

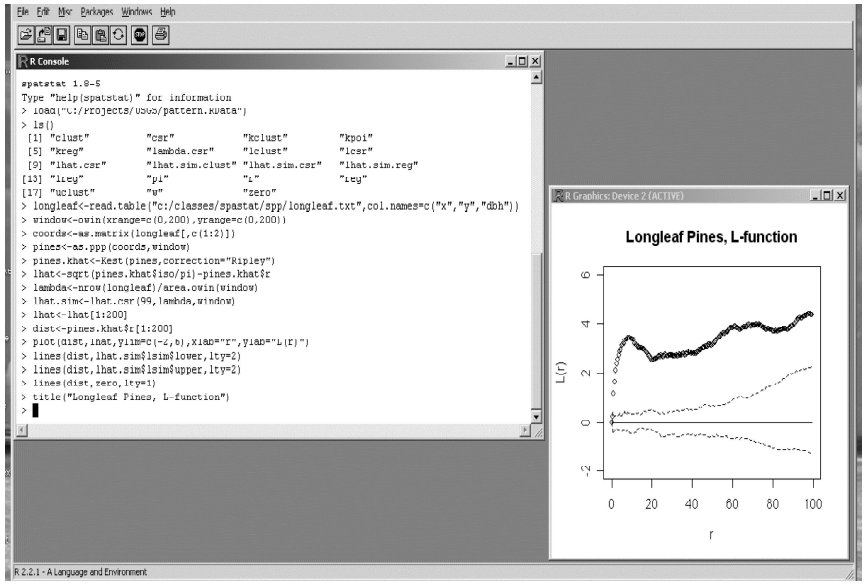
- spastat http://www.spatstat.org/

  Analysis of spatial point patterns.
- splancs http://www.maths.lancs.ac.uk/~rowlings/Splancs/

  Analysis of spatial and spatiotemporal point pattersn.
- MarkedPointProcess
  http://www2.hsu-hh.de/schlath/schlather.html#Software

  Analysis of marked point patterns.

R: spastat



Comments: R

- Public domain software;
- Packages contributed by statistical researchers keep the software up to date;
- Command driven and interactive;
- Spastat has procedures for:
  - Computing F-, G- and K-functions;
  - Testing complete spatial randomness;
  - Fitting the point cluster process model;
  - Simulating a variety of point process models;
  - Estimating parameters of modulated Poisson process model (covariates must be observed at all locations).

General Summary

- ArcGIS Geostatistical Analyst:
  - Menu driven;
  - A comprehensive collection of geostatistical methods;
  - Expensive.
- R:
  - Up-to-date methods for geostatistical and point pattern analyses;
  - Public domain;
  - Command driven and interactive.
- S+SPATIALSTATS:
  - Best for analysis of lattice data;
  - Menu driven.